Artificial Intelligence Lecture IV/V

Roshni Chakraborty

Dept. Of Information Technology

ABV-IIITM Gwalior

August 26, 2025

Outline I

Artificial Intelligence

Example

- Prompt *its raining now, I would like to eat*
- Response
  - Comfort Food Cravings
    - Ramen or pho – warm, slurpy, and satisfying
    - Grilled cheese tomato soup – classic and easy
    - Hot pot or shabu-shabu – if you're in the mood to cook at the table
    - Korean kimchi stew (kimchi jjigae) – spicy and warming
  - Snacky Lazy Mood
    - Instant noodles + egg + greens – quick upgrade to a classic
    - Popcorn or roasted nuts – great with a cozy movie
    - Mochi or warm pastries – if you've got a sweet tooth
  - Want me to help find something local (restaurant/delivery app suggestions)?
    - I can look it up for you — just let me know your city or area.
    - What kind of food are you in the mood for: something hot, spicy, sweet, or hearty?

Artificial Intelligence

Example

- Prompt *I live in India, its raining now, I would like to eat*
- Response
  - Classic Rainy Day Snacks
    - Pakoras (bhajiyas) – Onion, potato, or paneer pakoras with green chutney and chai
    - Vada Pav – Spicy, fried potato vada in a pav with garlic chutney
    - Samosas – Hot, crispy samosas with tamarind chutney
    - Corn on the cob (bhutta) – Roasted on fire with lemon, salt, and chili powder
  - Hearty Mains (if you're hungrier)
    - Khichdi – Light, warm, and perfect comfort food. Add a spoon of ghee!
    - Pav Bhaji – Spicy mashed veggies with buttered pav
    - Masala Maggi – Classic 2-minute Maggi with extra veggies or cheese
  - Hot Beverages:
    - Masala Chai – The ultimate rain companion
    - Filter Coffee – If you prefer South Indian flavors
- Would you like a quick recipe or want to order something online?

Artificial Intelligence

Example

Example III

Original Sentence "মেয়েটির পোশাক ছিল সাহসী, যা নিয়ে অনেকে প্রশ্ন তুলেছেন।"
(translated for Non Bengali reading audience)
Original Sentence Translated She wore provocative dresses, which raises questions among many
Label Biased
Type of Bias Framing Bias (specifically, Gender Bias)
Debiased Sentence "এই ঘটনার সঙ্গে ভুক্তভোগীর পোশাকের কোনো সম্পর্ক নেই — দোষী একমাত্র অপরাধীই।"
Debiased Sentence Translated This incident has no relationship about the victims clothes, the only culprit is the one who committed the crime

Artificial Intelligence

Definition and philosophy

- How to quantify *intelligence* ?
  - Are humans intelligent?
  - replicating human behavior is an early hallmark of intelligence
  - Are humans always intelligent?
  - Can non-human behavior be intelligent?
- Definition
  - *The study of how to make computers do things at which, at the moment, people are better*
  - *The branch of computer science that is concerned with the automation of intelligent behavior*
- Objectives
  - Systems that think like humans/rationally
  - Systems that act like humans /rationally

Artificial Intelligence

Thinking Humanly

- Cognitive Science
  - Very hard to understand how humans think
- Do we want a machine that beats humans in chess or a machine that thinks like humans while beating humans in chess?
  - Deep Blue supposedly DOESN'T think like humans
- Thinking like humans is important
  - Intelligent tutoring, Expressing emotions in interfaces

Artificial Intelligence

Thinking Humanly

- Laws of Thought
- Every art and every inquiry, and similarly every action and every pursuit is thought to aim at some good
  - If the human cannot tell whether the responses from the other side of a wall are coming from a human or computer, then the computer is intelligent.
- Rational behavior: doing the right thing
- Need not always be deliberative – Reflexive
- Reproducible, constructive or mathematically analyzable
- machines that act intelligent have to think intelligently too

Artificial Intelligence

Rational Agents

- An agent should strive to do the right thing
  - based on what it can perceive
  - and the actions it can perform
  - right action is the one that will cause the agent to be most successful
- Performance measure
  - An objective criterion for success of an agent's behavior
  - performance measure of a vacuum-cleaner
    - agent could be amount of dirt cleaned up
    - amount of time taken
    - amount of electricity consumed
    - amount of noise generated, etc.

Artificial Intelligence

Example

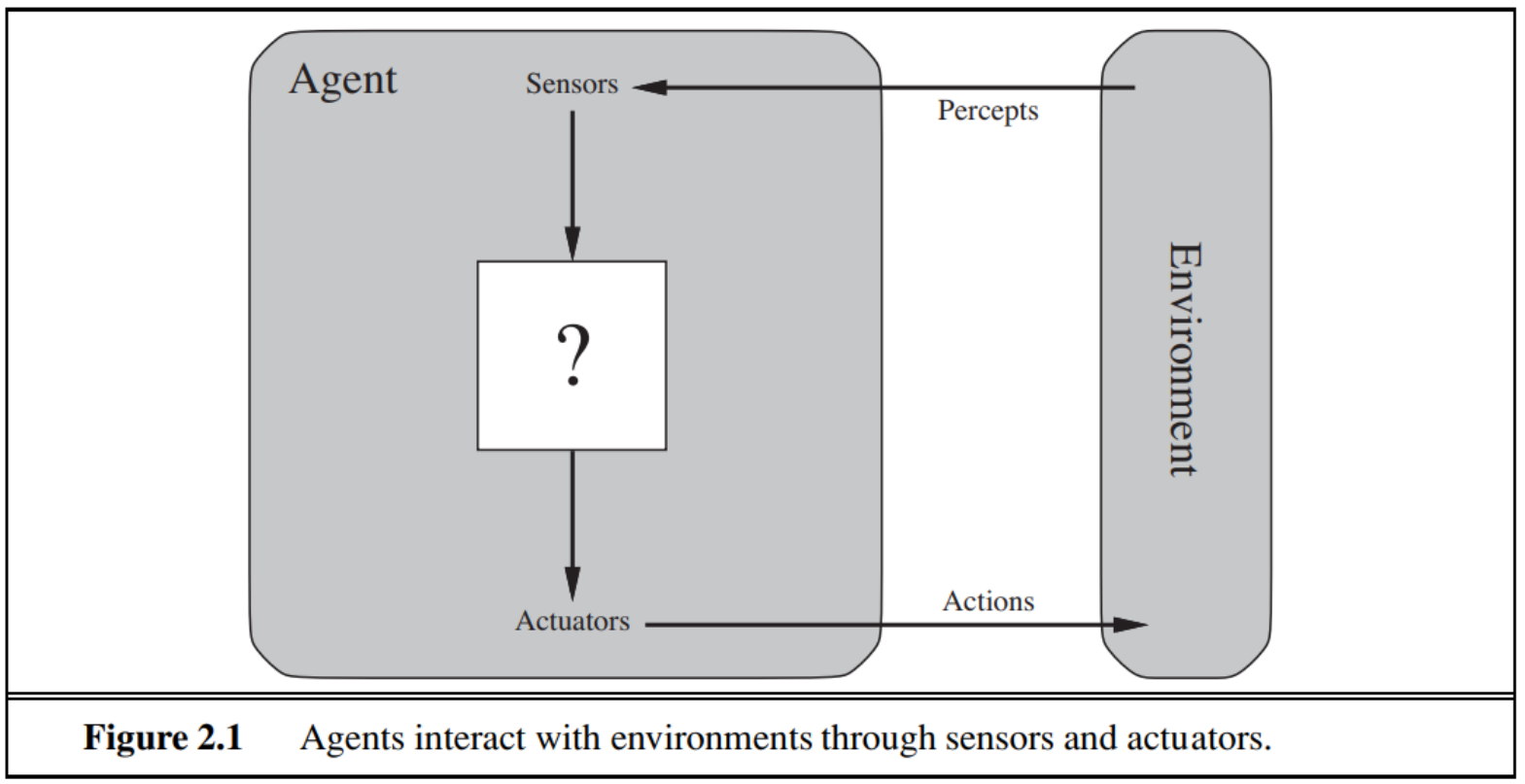

Figure 2.1 Agents interact with environments through sensors and actuators.

Artificial Intelligence

Ideal Rational Agent

- For each possible percept sequence,
  - does whatever action is expected to maximize its performance measure
  - on the basis of evidence perceived so far and built-in knowledge
- An agent is anything that can be viewed as
  - perceiving its environment through sensors
  - acting upon that environment through actuators
- Human agent
  - eyes, ears, and other organs for sensors
  - hands, legs, mouth, and other body parts for actuators
- Robotic agent
  - cameras and laser range finders for sensors– various motors for actuators
- **Objective** a good job of acting on their environment

Artificial Intelligence

Ideal Rational Agent

- This world is so simple that we can describe everything that happens
  - Its an imaginary world for our sake
- Environment
  - Two Squares A and B
- Perceive
  - which square am I in?
  - Is there dirt here?
- Actions
  - choose to move left, move right
  - cleanup or do nothing
- *if the current square is dirty, then clean; otherwise, move to the other square*

Artificial Intelligence

Percept Sequence

Percept sequence	Action
$[A, Clean]$	$Right$
$[A, Dirty]$	$Suck$
$[B, Clean]$	$Left$
$[B, Dirty]$	$Suck$
$[A, Clean], [A, Clean]$	$Right$
$[A, Clean], [A, Dirty]$	$Suck$
⋮	⋮
$[A, Clean], [A, Clean], [A, Clean]$	$Right$
$[A, Clean], [A, Clean], [A, Dirty]$	$Suck$
⋮	⋮

Figure 2.3 Partial tabulation of a simple agent function for the vacuum-cleaner world

Artificial Intelligence

Ideal Rational Agent

- A rational agent is one that does the right thing
- This is better than doing the wrong thing, implication ?
  - the right action is the one that will cause the agent to be most successful
  - decide how and when to evaluate the agent's success.
- performance measure the criteria that determine how successful an agent is
- considering the consequences of the agent's behavior
  - generates a sequence of actions according to the percepts it receives
  - this sequence of actions impacts the environment which undergoes a sequence of states
  - Is this *desirable*?
    - can be measured by *performance measure*

Artificial Intelligence

Ideal Rational Agent : Performance Measure

- Consider an agent which is vacuum a dirty floor
  - how about the *amount of dirt cleaned up in a single eight-hour shift*
  - how about the *amount of electricity consumed and the amount of noise generated*
  - how about the
    - cleans the floor quietly and efficiently
    - finds time to go windsurfing at the weekend.
- *clean floor* is based on average cleanliness over time
  - average cleanliness is better when?
    - one of which does a mediocre job all the time
    - other cleans energetically but takes long breaks

Artificial Intelligence

Ideal Rational Agent

- When to evaluate
  - measured how much dirt the agent had cleaned up in the first hour of the day
  - punishing those that work consistently
- The performance measure that defines degree of success
  - Everything that the agent has perceived so far
  - The percept sequence
  - What the agent knows about the environment
  - The actions that the agent can perform
- *For each possible percept sequence, an ideal rational agent should do whatever action is expected to maximize its performance measure, on the basis of the evidence provided by the percept sequence and whatever built-in knowledge the agent has*

Artificial Intelligence

Ideal Rational Agent

- Examples
  - Accident thing
  - Student life
  - Now apply to AI
- Rationality maximizes expected performance,
  - while perfection maximizes actual performance
- Our definition of rationality does not require *supernatural* powers
  - Also, does not allow *underintelligent activities*

Artificial Intelligence

Vaccum Cleaner Agent

- Objective cleans a square if it is dirty and moves to the other square if not
- Is this a rational agent?
  - what is the performance measure ?
    - awards one point for each clean square at each time step, over a *lifetime* of 1000 time steps
  - *geography* of the environment is known a priori
  - dirt distribution and the initial location of the agent are not
    - allowed actions are left, right and do nothing

Artificial Intelligence

Vaccum Cleaner Agent

- Can it ever be irrational
  - Keeps on oscillating across *clean* rooms
  - How to prevent?
    - a penalty of one point for each movement left or right
    - A better agent for this case would do nothing once it is sure that all the squares are clean
    - occasionally check and clean only if needed
- Two important factors for being rational agent are
  - information gathering and learning
  - the behavior of a rational agent can become effectively independent of its prior knowledge
    - enough built-in reflexes so that they can survive long enough to learn for themselves

Artificial Intelligence

Nature of Environment

- Definition of *rational agent* is clear
- Focus on building the *rational agent*
  - task environments are *problems*
  - rational agents are the *solutions*
  - Specifying the task environment
    - performance measure, the environment, and the agent's actuators and sensors (PEAS)
    - Design as much in detail as possible
- Example Automated Taxi Driver
  - performance measures
    - Reach correct destination
    - Minimize violations of traffic laws and disturbances to other drivers
    - Minimize fuel consumption and wear and tear
    - Maximize profile
  - Any conflicts in the goals? trade off required?

Artificial Intelligence

Nature of Environment

- Environment Issues
  - variety of roads, ranging from rural lanes and urban alleys to 12-lane freeways
  - stray animals, humans, other vehicles, different forms of vehicles, condition of roads
  - varies on the basis of location
- Actuators
  - accelerator, clutch, brake, steering wheel
- Sensors
  - controllable TV cameras, the speedometer, and the odometer, GPS
- Other Examples

Artificial Intelligence

Properties of Environment

- Fully Observable vs Partially Observable
  - A task environment is effectively fully observable if the sensors detect all aspects that are relevant to the choice of action
  - partially observable because of noisy and inaccurate sensors
  - automated drivers can not see what others are thinking
- Deterministic vs Stochastic
  - next state of the environment is completely determined by the current state and the action executed by the agent
  - an agent need not worry about uncertainty
  - Taxi driving is clearly stochastic in this sense
    - can never predict the behavior of traffic exactly
  - vacuum world as we described it is deterministic
    - what happens if smething suddenly happens?

Artificial Intelligence

Properties of Environment

- Episodic vs. sequentiaL
  - the agent's experience is divided into atomic episodes
  - Each episode consists of the agent perceiving and then performing a single action
  - independent of previous action
- Chess and taxi driving are sequential
- Static vs dynamic
  - If the environment can change while an agent is deliberating
  - environment is dynamic for that agent
  - Taxi driving is clearly dynamic
    - chess? puzzles?
  - Discrete vs. continuous
  - Single agent vs. multiagent

Artificial Intelligence

Properties of Environment

- EXAMPLE table

Artificial Intelligence

Simple Reflex Agent

- Select actions on the basis of the current percept
  - ignore the rest of the percept history
- Example Vaccum Cleaner
  - decision is based only on the current location
  - whether that contains dirt

Artificial Intelligence

Model based Agent

- To handle partial observability is
  - agent to keep track of the part of the world it can't see now
- Knowledge needs to be incorporated
  - information about how the world evolves independently of the agent
  - information about how the agent's own actions affect the world

Goal based Agent

- Prior knowledge and understanding of the goal
  - along with current state information
  - aids in understanding *desirability* of action
- goal-based agent appears less efficient
  - more flexible
  - knowledge that supports its decisions

Artificial Intelligence

Utility based Agent

- Goals alone are not always enough !
  - Need also to ensure high quality behavior
  - Example get the taxi to its destination *Goal* achieved
  - quicker, safer, more reliable, or cheaper
- A utility function maps a state (or a sequence of states) onto a real number
  - which describes the associated degree of happiness
  - For conflicting goals, utility function specifies the appropriate tradeoff

Learning Agents

- how the agent programs *come into being*
  - to operate in initially unknown environments
  - to become more competent than its initial knowledge alone might allow
- learning element which is responsible for making improvements
  - How am I going to get it to learn this?

Thank you!